

WHAT IS CLAIMED IS:

1. A method of predicting a biochemical network or pathway, the method comprising:

5 providing a population of compounds, which population comprises one or more input compounds and one or more output compounds;

defining at least one state-space that comprises the population of compounds; and,

10 identifying one or more candidate biochemical networks or pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space, thereby predicting the biochemical network or pathway.

2. The method of claim 1, wherein the informed search technique comprises one or more of: a greedy search, a uniform cost search, or an A\* search.

15 3. The method of claim 1, wherein the informed search technique comprises chemical distances between compounds in the population of compounds as an admissible heuristic.

20 4. The method of claim 3, comprising maintaining the admissibility of the heuristic by incorporating one or more penalties into an edge-cost associated with an edge in the state-space.

5. The method of claim 1, wherein the defining step comprises deriving one or more symbolic chemical substructural rules from biotransformations between two or more compounds in the population of compounds.

25 6. The method of claim 5, wherein the identifying step comprises recursively applying the symbolic chemical substructural rules to a selected compound to predict the biochemical network or pathway.

7. The method of claim 1, wherein the state-space is defined as two or more compound states and at least one state-transition between at least two of the compound states.

8. The method of claim 7, wherein at least one biocatalyst, at least one structural changes, and/or at least one energetic change describes the state-transition.

9. The method of claim 1, wherein at least one of the input and/or output compounds are described by one or more chemical descriptors.

10. The method of claim 8, wherein at least one of the chemical descriptors is selected from the list of descriptors provided in Figure 2.

11. The method of claim 8, wherein one or more of the input and/or output compounds is defined by at least one state vector.

12. A computer program product comprising a computer readable medium having one or more logic instructions for:  
15 receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds and one or more output compounds; and,  
identifying one or more candidate biochemical networks or pathways between at least one of the input compounds and at least one of the output compounds  
20 using at least one informed search technique to search the state-space.

13. The computer program product of claim 12, wherein the data comprises compound states and state-transitions between the compound states.

14. The computer program product of claim 12, wherein the computer readable medium comprises at least one logic instruction for:  
25 deriving one or more symbolic chemical substructural rules from biotransformations between two or more compounds in the population of compounds; and,  
recursively applying the symbolic chemical substructural rules to a selected compound to identify the candidate biochemical networks or pathways.

15. The computer program product of claim 12, wherein the computer readable medium comprises one or more of: a CD-ROM, a floppy disk, a tape, a flash memory device or component, a system memory device or component, a hard drive, or a data signal embodied in a carrier wave.

5           16. A system for predicting a biochemical network or pathway, comprising at least one computer having system software comprising one or more logic instructions for:

          receiving data that defines at least one state-space comprising a population of compounds, which population comprises one or more input compounds  
10       and one or more output compounds; and,

          identifying one or more candidate biochemical networks or pathways between at least one of the input compounds and at least one of the output compounds using at least one informed search technique to search the state-space.

15       17. The system of claim 16, wherein the data comprises compound states and state-transitions between the compound states.

          18. The system of claim 16, wherein the system software comprises at least one logic instruction for:

          deriving one or more symbolic chemical substructural rules from biotransformations between two or more compounds in the population of compounds;  
20       and,

          recursively applying the symbolic chemical substructural rules to a selected compound to identify the candidate biochemical networks or pathways.